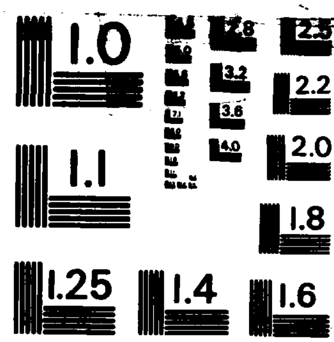
AD-A141 247 AN ESTIMATION PROBLEM IN FAULTY INSPECTION SAMPLING REVISED(U) MARYLAND UNIV COLLEGE PARK DEPT OF MANAGEMENT SCIENCES AND ST.. N L JOHNSON ET AL. JUN 84 1/1

UNCLASSIFIED UMD-DMSS-84/3 N00014-81-K-0301 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

③

AD-A141 247

An Estimation Problem in Faulty Inspection Sampling

by

Norman L. Johnson
University of North Carolina
Chapel Hill, N.C. 27514

Samuel Kotz
University of Maryland
College Park, Md. 20742

Revised Version

Key words and phrases: Faulty Inspection; Repeated Sampling; Binomial Distribution; Hypergeometric Distribution; Mixtures; Maximum Likelihood.

AMS 1982 subject classification 62N and 62P.

ABSTRACT

Suppose samples of size n are chosen at random from lots of size N and examined for the presence (or absence) of a certain type (or types) of defect. If inspection is not perfect, there may be a probability, p, of failing to detect a defect when it is, in fact, present. It is desired to estimate p from records of the number of items found to have the defect(s) on inspection.

In the absence of further information, the problem cannot be answered on the basis of single samples, of size n, from each of a number of lots. Some further assumptions are needed - such as supposing the number of defective items per lot is (nearly) constant. Alternatively, some more elaborate sampling - for example, repeated sampling from the same lot - may provide the necessary information.

The paper describes several methods of estimation utilizing various forms of additional data and information. Some as yet unsolved problems are briefly discussed.

I. Introduction

Suppose samples of size n are chosen at random from lots of size N and examined for the presence (or absence) of a certain type (or types) of defect. If inspection is not perfect, there may be a probability (1-p) of failing to detect a defect when it is, in fact, present. It is desired to estimate p from records of the number of items found to have the defect(s) on inspection. (We assume that the probability of 'false positive' - declaring a good item to be defective - is negligible.)

If sampling is with replacement, it is not possible to estimate p on the basis of single samples, of size n, from each of a number of lots, even supposing the actual proportion of defective items (θ) to be constant from lot to lot. This is because the distribution of the <u>observed</u> number of defectives (Z) in such a sample is binomial with parameters (n, θp) and it is not possible to disentangle p from θ.

If sampling is without replacement, again supposing the proportion of defectives, (θ = D/N when D is the number of defectives in the lot) the distribution of Z is a mixture of binomial distributions with parameters (Y,p) where Y - the actual number of defective items in the sample has a hypergeometric distribution with parameters (n,D,N). Conventionally we write (as in [1]).

$$Z \sim \text{Bin}(Y,p) \underset{Y}{\wedge} \text{Hypg}(n,D,N)$$

From a set of values of m independent random variables $(Z_1, Z_2, \ldots Z_m)$, knowing n and N, it is possible to estimate p by maximum likelihood from the likelihood function

$$L(Z_1, \ldots, Z_m) = \prod_{j=1}^{m} \left\{ \sum_{j=Z_j}^{n} \binom{y}{Z_j} p^{Z_j} (1-p)^{y-Z_j} \cdot \frac{\binom{D}{y}\binom{N-D}{n-y}}{\binom{N}{n}} \right\}$$

The limits of summation for y are $\max(Z_j, n-N+D) \leq y \leq \min(n,D)$. However, it is to be expected - in view of the fact that as N increases we approved the situation of sampling without replacement - that no great accuracy is estimating p will be attainable. A major part of difficulty in the nuisance parameters D(or $\theta = D/N$). For reasonable accuracy some form of repeated sampling will be needed.

2. Estimates of p

If we are in the fortunate position of having a number of items known to be defective we can obtain a simple estimator (of p) by testing them repeatedly and estimating p by the proportion of times a 'defective' decision is obtained. If this is not the case at the beginning of the investigation we might, however, be in a position to exploit the fact that an item declared 'defective' at any time must (according to our assumptions) be defective. Cenoting by N_j the number of items declared defective just j times in m trials (so that $N_0 + N_1 + \ldots + N_m = n$, and if $p = 1$, $N_0 = n-y$, $N_m = y$) a plausible but specious argument might run as follows:

> "For each item, we discard the first 'defective' decision and observe the proportion of defective decision in the remaining $\sum_{j=1}^{m} N_j$ sets of (m-1) trials. Since these are independent, the total number of defectives in the trials has a binomial $((m-1)\sum_{j=1}^{m} N_j, p)$ distribution, and our estimate of p is unbiased, with an easily computed standard deviation. (If $N_0 = m$, then no estimation is possible)."

(It is not difficult to see that this will produce a negatively biased estimator of p, because in all the trials which are thrown away a decision of 'defective' is reached. In fact the estimator is

$$\frac{1}{(m-1)\sum_{j=1}^{m} N_j}\left[\sum_{j=1}^{m}(j-1)N_j\right] = \frac{1}{m-1}\left[\frac{\sum_{j=1}^{m} jN_j}{\sum_{j=1}^{m} N_j} - 1\right] \tag{9}$$

and its expected value is

$$\left\{\frac{1}{m-1}\;\frac{mp}{1-(1-p)^m} - 1\right\} \cdot) \tag{10}$$

We could take notice of only those trials following the first 'defective' decision; although this will not use all the information available, it does lead to simple formulae. The observed proportion of 'defective' decisions is not an unbiased estimator of p; the (conditional) distribution of the number of 'defective' decisions counted is binomial with parameters (total number of inspections included in count, p).

Suppose, now, that we have m repeated inspections of a set of n items (among which an unknown number y are defective) and have been able to record the results of individual inspections (and not just total number of decisions of 'defective' for each inspection of the set of n items). The likelihood function is

$$\binom{y}{y-n+N_0, N_1, \ldots, N_m}(1-p)^{m(y-n+N_0)}\prod_{j=1}^{m}\left\{\binom{m}{j}p^j(1-p)^{m-j}\right\}^{N_j}$$

$$= \binom{y}{y-n+N_0, N_1, \ldots, N_m}\left\{\prod_{j=1}^{m}\binom{m}{j}\right\} p^{\sum_1^m jN_j}(1-p)^{my-\sum_1^m jN_j} \tag{11}$$

$$(n-N_0 \le y \le n)$$

where $N_0 = n - \sum_{j=1}^{m} N_j$ is the number of items which are not declared defective in any of the m inspections. Note that $(N_0, \sum_{j=1}^{m} jN_j)$ is a sufficient statistic for (y,p); $\sum jN_j$ is the total number of 'defective' decisions. If ~~y~~ were known, the maximum likelihood estimator of p would be

$$\hat{p}(y) = (my)^{-1} \sum_{ju}^{m} jN_j = (my)^{-1}T, \text{ say} \tag{12}$$

The corresponding maximized log likelihood would be

$$\log \hat{L}(y) = K + \sum_{i=0}^{n-N_0+1} \log (y-i) + (my-T) \log(my-T) - my \log my \tag{13}$$

where K does not depend on y. We then seek to maximize (13) with respect to y, subject to $n \geq y \geq n - N_0$. Note that we are not primarily interested in the value of y itself, but we need the value, y, maximizing (13) to calculate the maximum likelihood estimator, $\hat{p}(\hat{y})$, of p.

A useful practical method is obtained by noting that for the $\sum_{j=1}^{m} N_j$ items which we know to be defective, the numbers of times each item is declared 'defective' can be regarded as observed values of independent random variables each having a binomial (m,p) distribution, truncated by omission of zero values. Each such variable has expected value $mp\{1-(1-p)^m\}^{-1}$. Equating sample means and expected values gives the equation

$$m\tilde{p}\{1-(1-\tilde{p})^m\}^{-1} = (n-N_0)^{-1}T$$

= Average number of times an article is judged defective, given that it is so judged at least once.

This estimator is, in fact, the conditional maximum likelihood estimator of p, given N_0.

The equation can be solved iteratively, writing it in the form

$$\tilde{p} = m^{-1}(n-N_0)^{-1}T\{1-(1-\tilde{p})^m\} \tag{14}$$

We note that

$$\mathrm{var}(T) = mnp(1-p)\theta - m\, n(n-1)(N-1)^{-1}p^2\, \theta(1-\theta)$$

and

$$\mathrm{var}(\tilde{p}) \doteq m^{-2}\, (n-N_0)^{-2}\{1-(1-\tilde{p})^m\}^2\, \mathrm{var}(T)$$

If an estimator of y is needed, we note that, with

$$p = \tilde{p},\ E[\sum_{j=1}^{m} N_j = n - N_0 | y,\tilde{p}] = y\{1-(1-\tilde{p})^m\}.$$

Replacing expected by actual values, we get the estimator

$$\tilde{y} = \min(n, \left[(n-N_0)\{1-(1-p)^m\}^{-1}\right]) = \min(n, \left[(mp)^{-1}\, T\right]) \tag{15}$$

where [ ] denotes 'nearest integer to'.

As a numerical example suppose we test each of 50 (=n) items three (=m) times and obtain

$$N_0 = 43;\ N_1 = 1;\ N_2 = 1;\ N_3 = 5$$

so that

$$T = \sum_{j=1}^{3} jN_j = 1 + 2 + 15 = 18.$$ Equation (14) gives

$$\tilde{p}\{3\tilde{p} - 3\tilde{p}^2 + \tilde{p}^3\}^{-1} = (3 \times 7)^{-1}18 = 6/7$$

where $$\tilde{p}^2 - 3\tilde{p} + \frac{11}{6} = 0$$

leading to $$\tilde{p} = \underline{0.8545}$$

From (15), $\tilde{y} = \min\left[50, \left[\frac{18}{3 \times 0.8545}\right]\right] = 7.$

Note that the same values of $\tilde{p}$ and $\tilde{y}$ would be obtained, whatever the value of n (≥ 7).

Maximum likelihood estimation has been discussed by Lindsey [2].

ACKNOWLEDGEMENT

Samuel Kotz's work was supported by the U.S. Office of Naval Research under Contract N00014-81-K-0301. Norman L. Johnson's work was supported by the National Science Foundation under Grant MCS-8021704.

REFERENCES

([3]- contains background material in an earlier article by the authors).

[1] Johnson, N. L. and Kotz, S. (1984) Some Tests for Detection of Faulty Inspection. (Unpublished Report)

[2] Lindsey, B. G. (1983) Maximum Likelihood Estimation is an Error in Inspection Model Involving Sampling Without Replacement. Department of Statistics, Pennsylvania State University, Technical Report 44.

[3] Johnson, N. L. and Kotz, S. (1983) Faulty Inspection Distributions - Some Generalizations. In Reliability in the Acquisitions Process (D. J. De Priest and R. L. Launer, Eds.) pages 171-182. Dekker, New York.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. AD-A141 247

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
An Estimation Problem in Faulty Inspection Sampling

5. TYPE OF REPORT & PERIOD COVERED
TECHNICAL

6. PERFORMING ORG. REPORT NUMBER
UMD-DMSS-84/3

7. AUTHOR(s)
Norman L. Johnson and Samuel Kotz

8. CONTRACT OR GRANT NUMBER(s)
ONR Contract N00014-81-K-030
NSF Grant MCS-8-21704

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Management Sciences and Statistics
Univ. of Md., College Park, Md. 20742.

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
U.S. Office of Naval Research
Statistics and Probability Program (Code 436)

12. REPORT DATE
June 1984

13. NUMBER OF PAGES
8

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Faulty inspection; Repeated Sampling; Binomial Distribution; Hypergeometric Distribution; Mixtures; Maximum Likelihood.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Suppose samples of size n are chosen at random from lots of size N and examined for the presence (or absence) of a certain type(s) of defect. If inspection is not perfect, there may be a probability p, of failing to detect a defect when it is present. It is desired to estimate p from records of the number of items found to have the defect(s) on inspection. The paper describes several methods of estimation and briefly discusses some as yet unsolved problems.

END

FILMED

6-84

DTIC